Prosopography is Greek for Facebook: The SNAP:DRGN Project

Dr K Faith Lawrence
King's College, London
Dept. of Digital Humanities
26-29 Drury Lane, London WC2B 5RL
faith.lawrence@kcl.ac.uk

Dr Gabriel Bodard
King's College, London
Dept. of Digital Humanities
26-29 Drury Lane, London WC2B 5RL
gabriel.bodard@kcl.ac.uk

ABSTRACT
In this paper, we present SNAP:DRGN, a pilot project intended to support Ancient World Linked Open Data through the creation of persistent identifiers for person and person-like entities. We introduce the linked data landscape as it exists with respect to the digitized Classical world and SNAP:DRGN's place within it.

Categories and Subject Descriptors
H.3.5 [**Online Information Services**]: Data Sharing, Web-based Services I.2.4 [**Knowledge Representation Formalisms and Methods**]: Semantic Networks J.5 [**Arts and Humanities**]: Literature, Prosopography

General Terms
Management, Documentation, Standardization, Theory.

Keywords
Prosopography, open linked data, ancient history, Classics.

1. INTRODUCTION
The introduction of linked data techniques to historical data is becoming increasingly popular, in part supported by the cultural heritage institutions adoption of CIDOC CRM/FRBR-OO, or similar standards, for object description and the ongoing move to share that data. As museums and art galleries become a more integrated part of the Linked Open Data (LOD) web, so related research – especially within the digital humanities – can not only throw off the shackles of the much-despised 'digital silo' but can reuse identifiers generated by respected projects rather than merely creating project-specific identifiers and observing the 'linked' aspect of LOD more in spirit than in practice.

2. HISTORICAL DATA AND THE PEOPLE WHO USE IT
In 2014 the Institute for the Study of the Ancient World published a collection of reports from participants of the previous two years NEH-funded Linked Ancient World Data Institute[2]. The 30 reports reflect the range of disciplines and interests of the contributors, but also demonstrate the growing movement to bring the generated data together in a meaningful way. Taking the reports as a representative sample of the state of the domain, we can see the clear focus is on people, places, texts, and artifacts.

WebSci '15, June 28-July 01, 2015, Oxford, United Kingdom

ACM 978-1-4503-3672-7/15/06…$15.00
DOI: http://dx.doi.org/10.1145/2786451.2786496

Of these, only geodata is currently systematically catered for. The Pelagios: Enable Linked Ancient Geodata In Open Systems (PELAGIOS) project, created to assist in the introducing the potential of open linked data to historical geospatial references, is now in it's fifth year and third iteration. With 39 partners from 6 countries and over 830.000 annotations[3] the project represents one of the most well known and interconnected LOD projects in the sphere of the ancient world.

This successful model of a project consolidating historical geodata and place references will serve as an inspiration and invaluable collaborator for the largest desideratum: bringing together historical person references.

3. PROSOPOGRAPHICAL DATA
While social network platforms have brought into spotlight the advertiser and government appetite for the collation and classification of 'person' data in a scale previously unrivalled, the identification and correlation of person entities from disparate sources is a question that has long-plagued historians to the point of forming a distinct and recognized sub-discipline. Prosopography, as distinct from biography, focuses not on the life of an individual but on a person as part of a collection with whom they may share, or not, a number of traits. Pelteret[5] argues that "prosopography can be interpreted as the study of identifiable persons and their connections with others for the purpose of enabling the modern student to discern patterns of relationships."

The connection between prosopography and Linked Data is evident: people only become who they are through their relations with other entities, be they people, places or events. Emboldened by this conclusion, the Digital Classicist community started collecting together information about existing datasets, on- or offline, and their current state. The result, available at http://wiki.digitalclassicist.org/Greco-Roman_Prosopographies, identified twenty-four datasets of which seventeen were available online, wholly or in part. Of those, eight generated something resembling a persistent identifier or URI, and approximately four were available in some form of RDF.

For researchers who are becoming used to being able to annotate ancient places with established identifiers, the distribution for person URIs across unrelated projects places an added burden on encoding practices: providing a established URI requires first identifying which project (or projects) are the most appropriate to search for the entity and a subsequent the lack of clarity between that entity not having an established identifier and it having one but the entity being defined elsewhere if it is not immediately found.

3.1 SNAP: DRGN
The SNAP:DRGN (SNAP) project was created in response to this perceived gap in the coverage of ancient world linked entities. The Pleiades project, on which PELAGIOS builds, is able to relate their entries back to a single, respected source – the Barrington Atlas. Conversely the SNAP project had no single text which could act as canonical resource. Instead SNAP was faced with the results of multiple projects which had extracted data from a variety of potentially overlapping texts and other sources.

To create a stable starting point, SNAP partnered with three of the largest and most well known classical datasets: the Lexicon of Greek Personal Names, an Oxford-based corpus of persons mentioned in ancient Greek texts; Trismegistos, a Leuven-run database of names and persons from Egyptian papyri; Prosopographia Imperii Romani, a series of printed books listing senators and other elites from the first three centuries of the Roman Empire. The intention was not just to pilot a single lookup point for ancient people but to create a set of procedures and standards which would facilitate the addition of other datasets to the collection and allow the move from what Bradley calls 'closed' to 'open' prosopographies with "collaboration between partners, fuzzy boundaries, multiple overlapping interests"[1].

3.2 Defining the Model

The Friend of a Friend (FOAF) ontology offers a popular and lightweight model for describing person-data. In this emerged role as a *de facto* standard, FOAF has also been a first choice for many historical projects. However there has been a long-running discontent with the appropriation of a modern, and arguably western, model of a person for a historical entity whose fundamental existence is based on informed conjecture. The Factoid model[4], an alternative model for representing historical entities has also gained popularity, taking as it does, the assertions made in the primary source material as a base from which the person entity is formed. Other projects are content to identify persons as part of their TEI markup or CIDOC CRM encoding and from those structures extract an index.

It was decided that a lightweight model which could be mapped to the existing, competing, standards would allow the most flexibility and would fulfill the need of the project for a bridging model which would impose the minimum requirements on any potential collaborators. The model attempts to collect, and therefore duplicate from their originating projects, the minimal amount of data needed to aid disambiguation (see Figure 1).

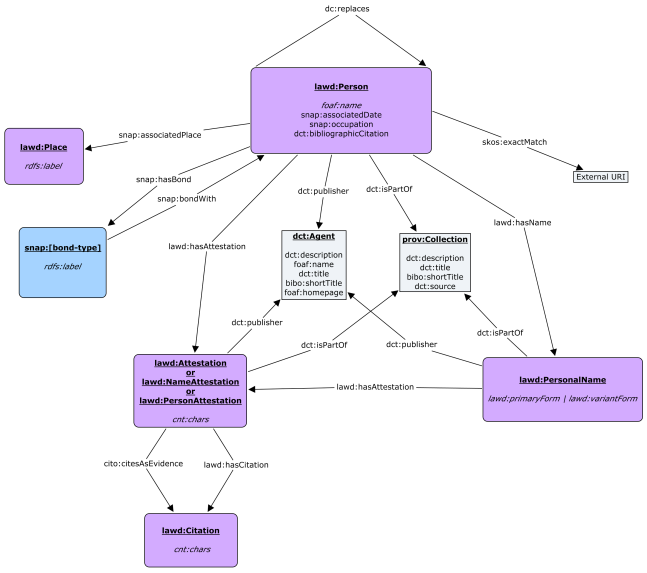

Figure 1: SNAP Data Model

4. DIRTY BY DESIGN

While modern prosopographical datasets have to contend with legal and ethical questions related to the collection of person-data, ancient prosopography faces different challenges. Although the identification of distinct entities to which a collection of qualified co-references refer is an important facet of prosopography, there is strong recognition that in many cases the conclusions can never be verifiably true. Therefore the aim of prosopographical projects is as much to present a scholarly statement of their truth as it is to present that truth. Part of that discussion means that disagreements exist, not to be resolved, because often that would be an impossibility, but to be exposed for consideration and future contemplation. The next iteration of the project will, in part, focus on the model required to express partial, contentious or theoretical merging (and dividing) of entities in line with scholarly practice.

5. CONCLUSION

The SNAP:DRGN was funded as a pilot project to investigate the potential of creating shared URIs for person, person-like and name entities in the Classical world by drawing together the disparate prosopographical datasets that already existed and to create a roadmap for new and upcoming prosopographical projects. At the time of writing 673753 person identifiers drawn from five projects have been publically released, providing standardized and persistent identifiers which link back to their originating projects[1].

In addition to the previously discussed model, a lightweight ontology[2] of interpersonal relationships designed with the needs of historical person entities has been developed and is being tested through application to further Classical world based projects and through mapping to Byzantine and later projects.

5. ACKNOWLEDGMENTS

Our thanks to the entire SNAP:DRGN team especially and our wonderful student volunteers.

6. REFERENCES

[1] Bradley, J. 'Towards an Ontology for Historical Persons'. Presented at 'Culturecloud, co-reference, Archive Workshop', National Archives, Stockholm, (2014).

[2] Elliott, T., Heath, S., Muccigrosso, J. (eds), 'Current Practice in Linked Open Data for the Ancient World', (2014). http://dlib.nyu.edu/awdl/isaw/isaw-papers/7/

[3] Isaksen, L., Barker, E., Simon, R. and de Soto, P. 2015. 'What Do You Do with a Million Links?' presented at 'Making Meaning from Data', AIA / SCS New Orleans (Jan. 2015). http://pelagios-project.blogspot.co.uk/2015/01/what-do-you-do-with-million-links.html

[4] Pasin, M.; Bradley, J., 'Factoid-based prosopography and computer ontologies: Towards an integrated approach'. In 'Literary and Linguistic Computing: the journal of digital scholarship in the humanities', (2013).

[5] Pelteret, D. ,'Unity in Diversity: Prosopographies and their Relationships with other Databases', History and Computing 12.1 (2000), pp. 13-22. http://www.wmich.edu/medieval/mip/journals/prosopog.htm

[1] http://snapdrgn.net/archives/321

[2] http://snapdrgn.net/ontology